



Technical note: Recursive rediscritisation of geo-scientific data in the Modular Earth Submodel System (MESSy)

P. Jöckel

► To cite this version:

P. Jöckel. Technical note: Recursive rediscritisation of geo-scientific data in the Modular Earth Submodel System (MESSy). *Atmospheric Chemistry and Physics Discussions*, 2006, 6 (3), pp.4673-4688. hal-00301518

HAL Id: hal-00301518

<https://hal.science/hal-00301518>

Submitted on 12 Jun 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MESSy NCREGRID

P. Jöckel

Technical note: Recursive rediscretisation of geo-scientific data in the Modular Earth Submodel System (MESSy)

P. Jöckel

Max Planck Institute for Chemistry, Air Chemistry Department, P.O. Box 3060, 55020 Mainz,
Germany

Received: 28 March 2006 – Accepted: 13 April 2006 – Published: 12 June 2006

Correspondence to: P. Jöckel (joeckel@mpch-mainz.mpg.de)

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

EGU

Abstract

This technical note describes a method for the rediscritisation of “gridded” geo-scientific data. A recursive algorithm (NCREGRID) is derived to solve the rediscritisation problem for orthogonal grids (also curvilinear) of arbitrary dimension. The algorithm is used within the program NCREGRID to handle geo-scientific data. These data are typically 2-dimensional (latitude-longitude grid), or 3-dimensional (latitude-longitude grid with a vertical pressure, or hybrid pressure coordinate, as used in atmospheric modelling). NCREGRID can be used as stand-alone program for the transformation (“re-gridding”) of data from and to data files in netCDF format. Moreover, NCREGRID constitutes the core of the Modular Earth Submodel System (MESSy) data import interface, providing a powerful tool for accessing data independently of the applied grid resolution (“automatic regriding”).

1 Introduction

In 3-dimensional global Earth system modelling the solving of sets of differential equations describing geophysical/chemical properties in the space time domain is often performed by discretisation following the Eulerian approach, i.e., on a discrete grid of spatial coordinates. Model input parameters need to be available on 2- or 3-dimensional (2-D, 3-D) grids of various resolutions. The horizontal grid space usually comprises geographical latitude and longitude. Especially in 3-D global atmospheric models the vertical pressure (p) coordinate is often defined by hybrid levels (with index i) of the form

$$p(i, x, y, t) = h_a(i) \cdot p_0 + h_b(i) \cdot p_s(x, y, t) , \quad (1)$$

where p_s is the surface pressure, p_0 is a constant reference pressure, and h_a and h_b are the dimensionless hybrid coefficients. This representation in a curvilinear coordinate system (dependent on longitude x , latitude y , and time t) allows a terrain following

ACPD

6, 4673–4688, 2006

MESSy NCREGRID

P. Jöckel

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

EGU

vertical coordinate, if $h_a=0$ and $h_b=1$ for the lowest level (surface level).

A wide variety of grid resolutions is used in different models and for different purposes. Moreover, global geophysical/chemical data (including observations) are provided in different spatial resolutions. As a consequence, data sets have often to be adapted to a specific grid resolution, i.e., rediscrctised, or “regridded”. For this purpose, NCREGRID has been developed.

The rediscrctisation is performed by a recursive algorithm (NREGRID), which is applicable to arbitrary orthogonal (also curvilinear) grids of any dimension, and not restricted to geo-hybrid-grid structures as described above. The algorithm does not apply a point-to-point interpolation, but a transformation based on overlaps between the different grid volumes.

NCREGRID can be used as a stand-alone program, and/or coupled as an import interface to a model, to rediscrctise automatically the input from an arbitrary grid space onto the required grid resolution. Specifically, NCREGRID provides the core of the Modular Earth Submodel System (MESSy, Jöckel et al., 2005) data import interface.

This technical note describes the construction of the algorithm (Sects. 2 and 3), the technical implementation (Sect. 4), the application of NCREGRID as MESSy data import interface (Sect. 5), and shows some examples (Sect. 6). More detailed information, user manuals, and the complete source code (of the current version 1.4b) with installation instructions can be found in the electronic supplement (<http://www.atmos-chem-phys-discuss.net/6/4673/2006/acpd-6-4673-2006-supplement.zip>). Future updates will become available on the internet¹.

2 The rediscrctisation problem

Let the function $F(\mathcal{A})$ be defined on a discrete grid $\mathcal{A} \subset \mathbb{R}^K$ of distinct grid intervals $A_n \subset \mathcal{A}$, where n is the index counting the grid intervals from 1 to N , K is the number of

¹<http://www.mpch-mainz.mpg.de/~joeckel/ncregid/>

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

dimensions, and \mathbb{R} the set of real numbers. The grid is defined such that

$$A_i \cap A_j = \emptyset \quad \forall i \neq j, \quad (2)$$

i.e., there is no overlap between different grid intervals. For the overall definition range (domain) of F it is

$$\mathcal{A} = \bigcup_{n=1}^N A_n. \quad (3)$$

The weighted average of F on its domain \mathcal{A} is

$$\bar{F} = \frac{\sum_{n=1}^N F_n \cdot |A_n|}{\sum_{n=1}^N |A_n|}, \quad (4)$$

where $|x|$ denotes the measure of x (e.g., the interval length, the area, or the volume in 1, 2, and 3 dimensions, respectively). F_n is the value of F within the grid interval A_n , i.e., $F_n := F(A_n)$ for simpler notation. Likewise, the unweighted sum of F is simply

$$\hat{F} = \sum_{n=1}^N F_n. \quad (5)$$

The function F should now be rediscritised onto a second grid \mathcal{B} consisting of M intervals, whereby \mathcal{B} has the same properties as \mathcal{A} above (Eq. 2). In particular, \mathcal{B} covers the same domain as \mathcal{A} , i.e.,

$$\mathcal{B} = \bigcup_{m=1}^M B_m = \bigcup_{n=1}^N A_n = \mathcal{A}. \quad (6)$$

With the measure introduced above, two different overlap matrices \mathbf{P} and \mathbf{Q} can be defined:

$$P_{n,m} := \frac{|A_n \cap B_m|}{|A_n|} \quad (7)$$

and

$$Q_{n,m} := \frac{|A_n \cap B_m|}{|B_m|}, \quad (8)$$

resulting in two different rediscrctised distributions of F on \mathcal{B} , denoted as G^E and G^I , respectively:

$$G_m^E := \sum_{n=1}^N P_{n,m} \cdot F_n = \sum_{n=1}^N \frac{|A_n \cap B_m|}{|A_n|} F_n \quad (9)$$

and

$$G_m^I := \sum_{n=1}^N Q_{n,m} \cdot F_n = \sum_{n=1}^N \frac{|A_n \cap B_m|}{|B_m|} F_n. \quad (10)$$

For further considerations, the general relation

$$|x \cup y| = |x| + |y| - |x \cap y| \quad (11)$$

can be expanded for a summation, yielding

$$\begin{aligned} \sum_{m=1}^M |A_n \cap B_m| &= \left| \bigcup_{m=1}^M (A_n \cap B_m) \right| \\ &= |A_n \cap (\bigcup_{m=1}^M B_m)| \\ &= |A_n \cap \mathcal{B}| \end{aligned} \quad (12)$$

The second conversion is applicable because of Eq. (2). Since furthermore \mathcal{A} and \mathcal{B} share the same domain by definition (Eq. 6), it further holds

$$A_n \subseteq \mathcal{B} \Rightarrow A_n \cap \mathcal{B} = A_n \quad (13)$$

and therefore

$$\sum_{m=1}^M |A_n \cap B_m| = |A_n| . \quad (14)$$

Finally, using Eq. (14), for the overlap matrices P and Q it is

$$\sum_{m=1}^M P_{n,m} = \sum_{n=1}^N Q_{n,m} = 1 . \quad (15)$$

- 5 With the definitions above and relation (14) it is straightforward to show for the un-weighted sum of G^E (cf. Eq. 5) that

$$\begin{aligned} \hat{G}^E &= \sum_{m=1}^M G_m^E \\ &= \sum_{m=1}^M \sum_{n=1}^N \frac{|A_n \cap B_m|}{|A_n|} F_n \\ &= \sum_{n=1}^N \frac{F_n}{|A_n|} \cdot \sum_{m=1}^M |A_n \cap B_m| \\ &= \sum_{n=1}^N F_n = \hat{F} . \end{aligned} \quad (16)$$

Similarly, the weighted average of G^I yields (cf. Eq. 4)

$$\begin{aligned} \bar{G}^I &= \frac{\sum_{m=1}^M G_m^I |B_m|}{\sum_{m=1}^M |B_m|} \\ &= \frac{\sum_{m=1}^M \sum_{n=1}^N |A_n \cap B_m| F_n}{\sum_{m=1}^M |B_m|} \end{aligned} \quad (17)$$

$$= \frac{\sum_{n=1}^N |A_n| F_n}{\sum_{n=1}^N |A_n|} = \bar{F}.$$

Note that in the denominator Eq. (6) has been used.

The choice of the overlap matrix (**P** or **Q**) for the grid transformation needs to consider which moment of the original discretised function F should be conserved: **P** conserves the unweighted sum and is therefore suitable for the rediscritisation of extensive variables (EXT, therefore the superscript E of G^E), whereas **Q** conserves the weighted average, suitable for the rediscritisation of intensive variables (INT, therefore the superscript I of G^I). Intensive quantities are defined at each point of the interval and do not depend on the interval size, whereas extensive quantities are representative for the interval as it is defined, and therefore scale with the interval size. Examples for intensive quantities are temperature, tracer mixing ratios, or an average altitude. In contrast, absolute tracer masses are extensive. Figure 1 illustrates the rediscritisation of INT and EXT quantities.

Finally, it should be noted that the algorithm can also be used to rediscritise non-continuous functions F , which can only take discrete values:

$$F \in \{f_1, f_2, \dots, f_u\} \quad (18)$$

The non-fractional land-sea mask with $u=2$, $f_1=1$ (land), and $f_2=0$ (ocean) is a simple example. For these “index functions”, averaging over intervals is not necessarily defined. However, with the definition

$$\tilde{F}[v] = \begin{cases} 1 & : F = f_v \\ 0 & : F \neq f_v \end{cases} \quad \forall v \in \{1, \dots, u\} \quad (19)$$

a rediscritisation is still possible, by first transforming all $\tilde{F}[v]$ as intensive quantities (INT), i.e., by calculating $\tilde{G}'_m[v]$ using the overlap matrix **Q**. The result $\tilde{G}'_m[v] \in [0, 1]$ is the fraction of index value f_v in grid box m of the destination grid. If a single index value

is desired, it can for instance be defined as the index with the highest fraction, i.e.,

$$\tilde{G}_m^I = f_w \ ; \ \tilde{G}_m^I[w] := \max_v \tilde{G}_m^I[v] \ . \quad (20)$$

3 The recursive algorithm

So far, no assumptions on the specific form or dimension of the discretisation intervals $A_n \subseteq \mathcal{A}$ and $B_m \subseteq \mathcal{B}$ have been made. Only the more general constraints of distinct intervals (Eq. 2) and of full coverage (Eq. 6) have been taken into account.

In practice, the discretisation domain in \mathcal{K} dimensions often describes a grid which is spanned by orthogonal axes. In these cases, an efficient recursive algorithm for calculation of the overlap matrices can be derived.

The scalar indices n or m (from 1 to N and M , respectively), for the unambiguous enumeration of the abstract discretisation intervals, expand to vector indices as

$$n \rightarrow \mathbf{n} = (n_1, n_2, \dots, n_{\mathcal{K}}) \ , \quad (21)$$

and likewise for m , running from $(1, 1, \dots, 1)$ to $(N_1, N_2, \dots, N_{\mathcal{K}})$ and $(M_1, M_2, \dots, M_{\mathcal{K}})$, respectively.

With this, the overlap matrices can be decomposed into \mathcal{K} sub-matrices of the form

$$P_{n,m} \rightarrow P_{\mathbf{n},\mathbf{m}} = P_{n_1,m_1} \bullet P_{n_2,m_2} \bullet \dots \bullet P_{n_{\mathcal{K}},m_{\mathcal{K}}} \quad (22)$$

(and likewise for Q), whereby the bullet (\bullet) indicates the element-wise multiplication of the matrix elements. It is therefore clear that an overlap matrix element $P_{\mathbf{n},\mathbf{m}}$ is zero, if the matrix element of at least one sub-matrix $1 \leq i \leq \mathcal{K}$ is zero, i.e., if $P_{n_i,m_i} = 0$.

Using this information, a recursive algorithm can be designed which calculates the overlap matrix for the first dimension, i.e., for P_{n_1,m_1} . For all elements where P_{n_1,m_1} is zero, also $P_{\mathbf{n},\mathbf{m}} = 0$ and the calculation is finished. For those elements where $P_{n_1,m_1} \neq 0$, the algorithm is called to calculate the overlap sub-matrix for the second dimension, i.e., P_{n_2,m_2} , and so on.

Furthermore, the dimensions can be ordered such that potentially invariant dimensions, defined by

$$P_{n_i, m_i} = \begin{cases} 1 & : n_i = m_i \\ 0 & : n_i \neq m_i \end{cases} \quad (23)$$

(i.e., where P_{n_i, m_i} is the identity matrix), are skipped at the end of the computation.

4 Technical implementation

NCREGRID is coded according to the Fortran95 standard (ISO/IEC-1539-1). The code quality has been tested by applying the Fortran analyser forcheck². The program can be applied in two different modes:

- As a stand-alone program (stand-alone mode) for the rediscretisation of data files. In this case, the grid information of the destination grid is imported from an additional data file.
- Coupled to a model (interface mode, e.g., linked to a Chemistry Transport Model (CTM), a General Circulation Model (GCM), etc.) for the resolution independent import of data from external files (automatic regridding). In this case, the grid information of the destination grid (= model grid) is provided by a model-specific interface routine.

In both modes, NCREGRID is controlled via the standard Fortran95 user interface (namelists). The user specifies the input data file, the name of the output data file, and describes the grid information of the input file. In the stand-alone mode additionally the destination grid information has to be provided.

The program is hierarchically organised into the following modules:

²<http://www.forcheck.nl>

1. BASE: This module contains NREGRID, which is the implementation of the recursive algorithm described in Sect. 3 for an arbitrary number \mathcal{K} of orthogonal dimensions. To allow an arbitrary number of dimensions, all arrays are internally transformed to arrays of rank 1 with index n by an unambiguous mapping procedure

$$(n_1, n_2, \dots, n_{\mathcal{K}}) \leftrightarrow n. \quad (24)$$

This is required, since Fortran95 does not allow the specification of the rank of an array during runtime.

2. NETCDF: This module organises the input and output from and to the external data files. The data file format for gridded data is netCDF³.
3. GEOHYB: This module meets the specific requirements of geo-hybrid grids (spherical geographic coordinates, vertical curvilinear pressure or hybrid pressure coordinates), and transforms the information (for instance including adequate weights for the latitude axis) for subsequent application of NREGRID.
4. DIAG: This module contains diagnostic routines which are only required for debugging.
5. CONTROL: This module contains the sequence control system of the overall re-gridding procedure: reading the user information, input data and grid information, reading output grid information, re-gridding, output of data. Since it is the main entry point for using NCREGRID, this module comprises the low-level (level 1) user-interface for the usage of NCREGRID coupled to a model. In the stand-alone mode, the re-gridding procedure is called via this level-1 user interface from the main program.

³<http://www.unidata.ucar.edu/software/netcdf/>

6. **TOOLS:** This module contains high-level (level 2) interface routines for the application of NCREGRID in interface mode.
7. **INTERFACE:** This module defines the grid information, if NCREGRID is coupled to a model (see Fig. 2). In the stand-alone mode, the grid information is imported from a data file; in this case this module is empty.

5

10

Installation instructions for the stand-alone mode, details about the namelist control, information for using the interface-mode (including the level 1 and level 2 interfaces), more details about the implementation, and the complete source code (of current version 1.4b) are contained in the electronic supplement (<http://www.atmos-chem-phys-discuss.net/6/4673/2006/acpd-6-4673-2006-supplement.zip>).

5 The MESSy data import interface

15

The coding of NCREGRID summarised in Sect. 4 strictly adheres to the MESSy standard (Jöckel et al., 2005). Specifically, the modules clearly separate into the submodel core layer (SMCL), namely BASE, NETCDF, GEOHYB, and DIAG, and into the submodel interface layer (SMIL), namely CONTROL, and TOOLS. The stand-alone mode with the main program of NCREGRID applies the level 1 interface (CONTROL) and constitutes a MESSy box model.

20

Moreover, NCREGRID is a generic MESSy submodel, since it provides (in interface mode) an essential part of the MESSy infrastructure, namely a general data import interface (for gridded data) with an automatic regridding facility. Technically, this is achieved by two additional modules for the base model interface layer (BMIL) organising the communication with the base model (e.g., a general circulation model (GCM)). These two additional BMIL modules are

25

- the module INTERFACE providing the base model grid information to NCREGRID, and

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

- a module providing the level 3 interface for passing the imported data from NCRE-GRID back to the base model.

As defined in the MESSy standard, the BMIL takes into account the coding aspects of the base model. Specifically, the level 3 interface routines need to transform the output of NCREGRID (Fortran95 structures) into the data fields of the base model, taking into account, for instance, the domain decomposition in a parallel model, the configuration and order of the dimensions, etc. In doing so, the level 3 module makes use of the level 1 and 2 routines of the SMCL. Figure 2 summarises the structural design aspects of NCREGRID in stand-alone and interface mode. The documentation of the BMIL for the the atmospheric chemistry GCM ECHAM5/MESSy (version 1.1) is provided in the supplement (<http://www.atmos-chem-phys-discuss.net/6/4673/2006/acpd-6-4673-2006-supplement.zip>).

6 Examples

Two example results of the application of NCREGRID are shown in Fig. 3. The upper left panel shows the Earth topography on a regular 2° by 2° latitude–longitude grid. The colour code represents the orography in km. The result after rediscretisation of this intensive quantity onto a regular 10° by 10° grid is shown in the upper right panel.

The lower left panel depicts a typical vertical distribution of an atmospheric tracer. A cross section along one longitude of a 3-dimensional dataset is shown. The quantity (colour code) is given as mixing ratio in arbitrary normalised units. In the vertical direction, the original data are on constant pressure levels (in hPa). The lower right panel shows the result (at the same longitude) after rediscretisation onto a different grid. The latitudinal, longitudinal, and the vertical grid intervals changed, specifically, the vertical axis of the destination grid is defined in hybrid pressure coordinates (Eq. 1). Note that this is not visible in the figure, since a constant surface pressure has been assumed for plotting. The black crosses in the lower panels indicate the mid points of the boxes

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

of the underlying grids.

7 Conclusions

This technical note describes a recursive algorithm and its implementation (NCRE-GRID) for the rediscritisation of (“gridded”) geo-scientific data from one grid to another. In contrast to point-to-point interpolation, the present algorithm is unambiguous and conserves integral moments, such as the weighted average of intensive quantities, and the unweighted sum of extensive quantities. No arbitrary information is added if data are rediscritised from a coarser to a finer grid, and vice versa, only minimal information is lost. Implemented in a modular structure following the MESSy standard, NCREGRID constitutes a generic MESSy submodel: NCREGRID itself is structured as a MESSy submodel, i.e., separated in submodel core layer and submodel interface layer, including a MESSy box-model (i.e., the stand-alone mode). With two additional modules in the base model interface layer, NCREGRID serves at the same time as the overall MESSy data import interface for automatic regridding of imported gridded data.

Acknowledgements. Helpful comments by R. Sander, A. Kerkweg and H. Tost are gratefully acknowledged. The author wishes to acknowledge use of the Ferret program for the graphics in this paper. Ferret is a product of NOAA’s Pacific Marine Environmental Laboratory (Information is available at <http://www.ferret.noaa.gov>).

References

Jöckel, P., Sander, R., Kerkweg, A., Tost, H., and Lelieveld, J.: Technical note: The Modular Earth Submodel System (MESSy) – a new approach towards Earth System Modeling, Atmos. Chem. Phys., 5, 433–444, 2005. [4675](#), [4683](#)

MESSy NCREGRID

P. Jöckel

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

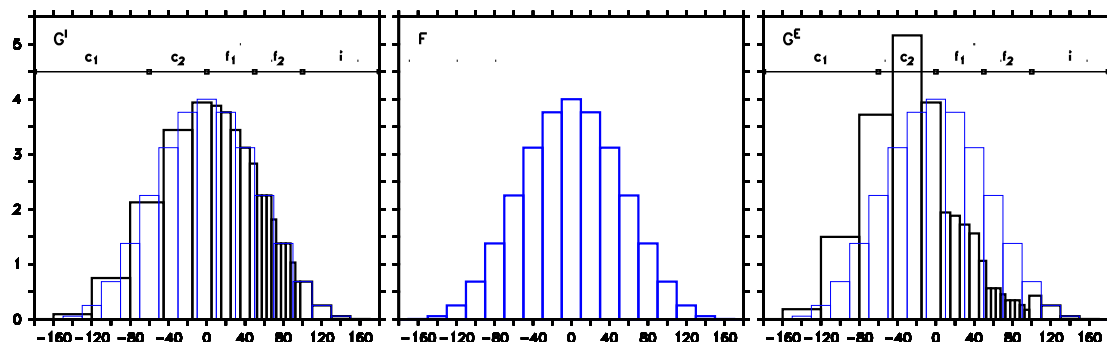


Fig. 1. Sketch of the rediscretisation procedure for a function F (blue, middle panel) in 1 dimension (horizontal axis). The original grid is regular with equidistant intervals. The destination grid is coarser within the intervals indicated by c_1 , and c_2 , and finer in the intervals f_1 , and f_2 , respectively. In interval i , destination and original grid are the same (invariant). The left panel shows the result (black), if F describes an intensive quantity (G^I , see text), and the right panel shows the result (black) for an extensive quantity (G^E , see text).

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[◀](#)
[▶](#)
[◀](#)
[▶](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)

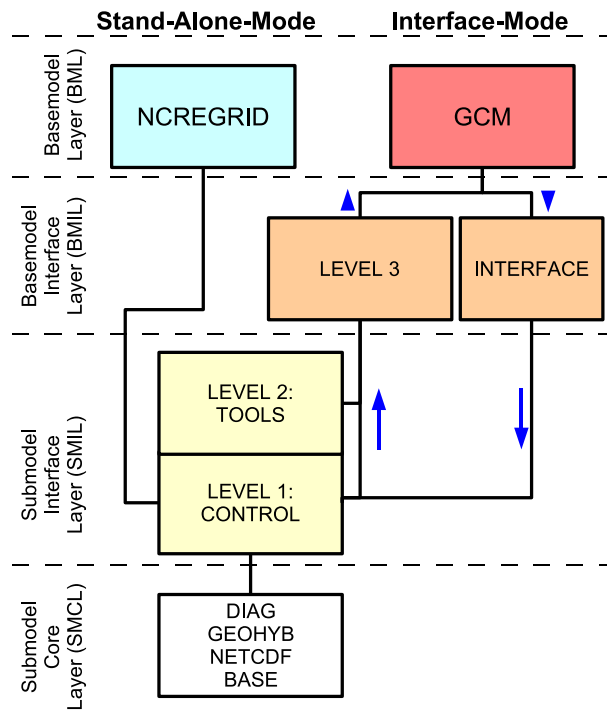


Fig. 2. Modular structure of NCREGRID and the MESSy data import interface. The modules in the submodel core layer (SMCL) and submodel interface layer (SMIL) are described in Sect. 4. For the stand-alone mode, NCREGRID provides the main program (constituting a MESSy “box” model). The alternative interface mode (here indicated by the coupling to a general circulation model (GCM) in the base model layer (BML)) requires the base model grid description to be provided in the module INTERFACE, and an additional module with level 3 interface routines, providing the base model interface layer (BMIL). The level 3 routines need to take into account the specific design aspects of the base model, here the GCM. The blue arrows indicate the information flow.

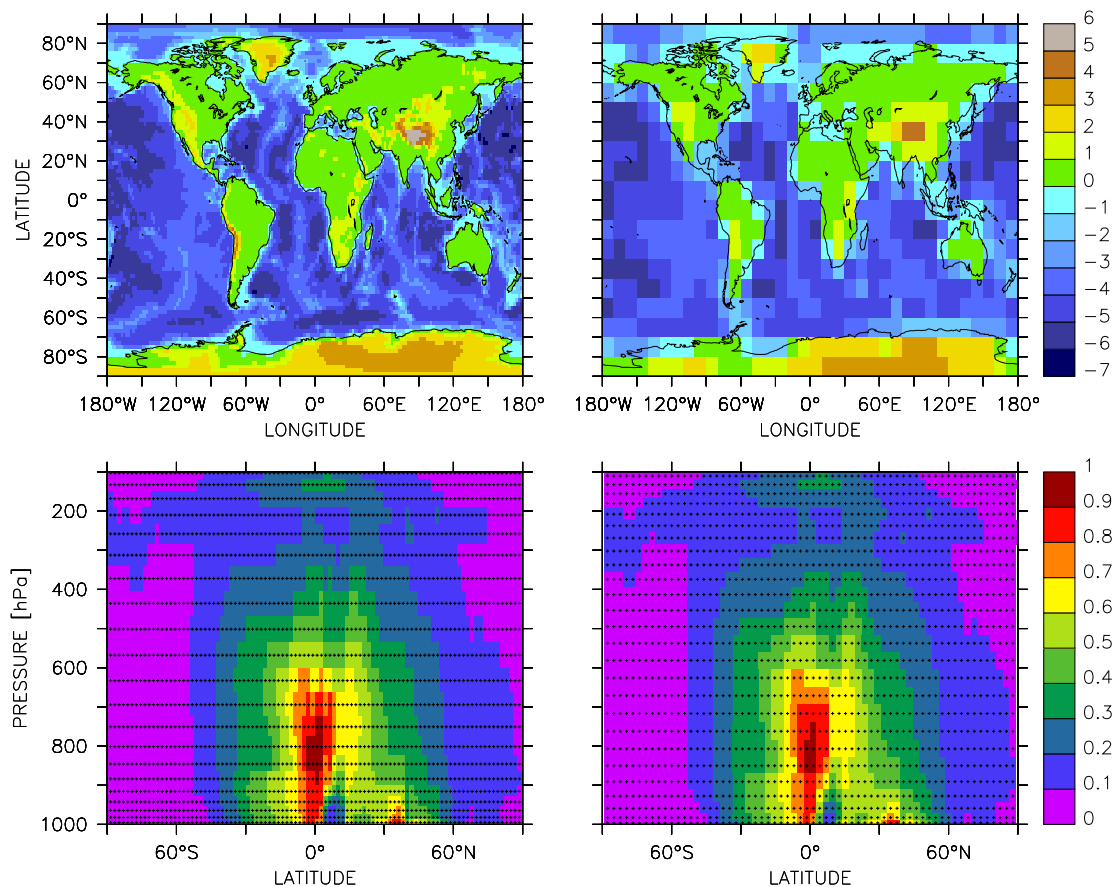


Fig. 3. Examples for the application of NCREGRID showing topographical height in km (top) and a vertical tracer distribution in normalized units (bottom). Details are given in the text.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)